CS70 Discussion 8b (CLT Review)

Agnibho Roy

Summer 2020

1 Penguins!

Professor Sahai decides that he wants to vacation but wants to do so in isolation due to the coronavirus so he ventures to Antarctica. He read once that penguins have a height anywhere from 3ft to 5ft with uniform probability, but is skeptical so decides to see for himself. We want to see how closely the average of the penguins he measures is to the true average. Let:

$$\hat{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

be the average of the n height samples from the population of penguins, each independently and randomly collected.

1. Calculate the expected value and variance for the height of a penguin

2. Calculate a 95% confidence interval for the average height of the penguins for an arbitrary n using Chebyshev's Inequality. Interpret this interval.

3. Calculate a 95% confidence interval for the average height of the penguins for an arbitrary n using CLT. You may assume that n is sufficiently large. You may assume that $Pr(-1.96 < \mathcal{N}(0,1) < 1.96) = 0.95$

4. Which of the methods provides a tighter interval for any value of n? What additional conditions do you need to be able to use CLT?

2 Blood Pressure

We have a random variable Y denoting the blood pressure of a patient, and suppose we model it as a Gaussian random variable having mean μ_y, which we know for a fact that $\mu_y \in [60, 90]$, and variance $\sigma_y^2 = 12$. The blood pressure monitor is faulty and has a measurement $X = Y + W$ where W is Gaussian with zero mean ($\mu_w = 0$) and variance $\sigma_w^2 = 4$ (uncorrelated to Y). We estimate the true mean of Y using the estimator:

$$\hat{\mu}_y = \frac{X_1 + \cdots + X_n}{n}$$

Where X_i are independent measurements of the random variable $X = Y + W$. We want to find the minimum number of measurements such that $|\hat{\mu}_y - \mu_y|$ is within 4% of μ_y. (CS 70 SP18 Final)

3 Binomial Concentration

Here, we will prove that the binomial distribution is concentrated about its mean as the number of trials tend to ∞. Suppose we have i.i.d trials, each with a probability of success $\frac{1}{2}$. Let S_n be the number of successes in the first n trails (n is a positive integer), and define:

$$Z_n := \frac{S_n - n/2}{\sqrt{n}/2}$$

1. What are the mean and variance of Z_n?

2. What is the distribution of Z_n as $n \to \infty$?

3. Use the bound $P(Z > z) \leq (\sqrt{2\pi}z)^{-1}e^{-z^2/2}$ when Z is a standard normal in order to bound $P(\frac{S_n}{n} > \frac{1}{2} + \delta)$, where $\delta > 0$.